

Rapid Data Access: Key to Integrated Use of Environmental Characterization and Monitoring Information

Thomas Canales, Patricia Ottesen,
Lawrence Livermore National Laboratory

Abstract:

Environmental investigations result in large quantities of data. The value of these data lies in their interpretation and use by project staff, management, and the regulatory community. Traditional modes of data access can be frustrating and time-consuming. Software tools that join networking technology of the World Wide Web (WWW) with database access have decreased labor-intensive overhead in site characterization and monitoring, thus increasing the efficiency of the ground water restoration project at Lawrence Livermore National Laboratory (LLNL).

DOE and project personnel have dynamic access to statistical processing, database retrieval, and cost estimating tools. By adding mouse-sensitive site maps and post-processing capabilities, we have extended the utility of standard web browsers, such as Netscape or Mosaic. Users can retrieve chemical compound concentrations, ground water elevations, descriptive information about monitoring locations, or analytical services' cost estimations. Data may be viewed as time-series graphs, contour maps, chemical concentration histograms, or simply displayed as text. Platform independence and easy retrieval make more comprehensive review of data possible. Cost-savings are realized; efficiency is increased; planning and decision-making are facilitated.

Extended Summary:

LLNL uses a dedicated, customer service team approach to information management. Interdisciplinary professionals and technicians facilitate clear communication with the varied project staff. Traditional data access modes required the intervention or assistance of information management team members for project staff to obtain data for analysis and interpretation. This close working relationship enabled the information management team to become familiar with recurring and frequent data requests. Storage of data is designed based on projected retrieval needs. The value of the data is in their use, rather than simply secure storage. The team participates in producing project deliverables and in meeting deadlines, rather than retroactive archive activities, although an excellent archive is also achieved.

Data elements needed by each discipline are captured to provide a consistent data set that is available to all. Data quality is identified and stored electronically. Data of known quality may be used at later dates. Single-entry and multiple use of data are cost-effective information management strategies. Reduced data from one discipline are electronically available for retrieval and use by others. Validated and qualified chemistry data are

available for geologists, hydrologists, and engineers. Hydrostratigraphic unit identification is available for chemists, engineers, and modelers. The retrieval and integrated use of correlated data are enhanced by the use of a location-based information model. This successful system enables project personnel to have timely and cost-effective access to current and historical data.

The dedicated customer service information management team assists project staff in obtaining data in formats appropriate to the intended use of the data. Standard format tabular reports are produced for documents and appendices. Custom reports are created to meet unique customer specifications. Ad hoc questions are answered rapidly either verbally or with a listing included in an e-mail message. Direct and timely access by project staff for common data requests is achieved by integrating the WWW's platform independent hypertext technology with X-window applications and traditional access methods of relational database management systems.

The WWW is implemented using a client-server model. Resources (text, images, sounds) are downloaded from a machine on the Internet (a server) to your local machine (a client) and viewed with a web browser such as Netscape or Mosaic. Each resource has its own unique address on the WWW called a Universal Resource Locator (URL). This URL is composed of a unique machine name and location on the machine where the resource can be found. Served resources can themselves contain references to other URLs. These hyperlinks are mouse-sensitive regions of text that, when selected, cause a request to be sent over the Internet to the server machine addressed in the URL asking that the resource be downloaded. The recent explosive growth of the WWW gives some indication of its power and utility in making a wide variety of information easily available to a large audience.

The X-window system is a mature client-server windowing system developed at MIT. It specifies a high level communications protocol between two programs: a server that performs drawing on your screen and handles input from your mouse and keyboard; and a client that performs the computational work and produces the graphical output. The power of the X-window system is that the two programs can reside on the same machine or on different machines. When an X-client application program is run on one system and displayed on another it is referred to as a remote X-client. Although the X-window system was initially developed for UNIX workstations, currently many vendors provide X-servers for the Mac and PC desktop community. Today much of the scientific X-application software previously available only to the UNIX workstation community can be made available to these other platforms.

Through the use of embedded relational database management system (RDBMS) languages like Structured Query Language (SQL) and form-based graphical user interfaces provided by RDBMS vendors, sophisticated software tools can be constructed that provide users with task specific products. Custom reports and tabular data files can be produced by users with little

knowledge of the underlying information model or query language. This functionality can be delivered and even extended by using the WWW. Emulating the vendor supplied form-based query tool with HTML forms is well understood and widely used. Typically, an HTML form is constructed with entry fields that allow the user to constrain a broad retrieval request to a specific query. As an example, consider a form that allows a user to extract time series data of a given parameter for a specified well. A web page containing an entry field for the well identifier and a menu listing of parameters (ground water elevation, extraction flow volume, contaminant concentration) could be created. A user provides a unique identifier indicating the well of interest and makes selections from the parameter list. When submitted, the form passes the well identifier and list of selected parameters back to a script on the web server, which in turn passes the selections to an embedded SQL program designed to process and submit a query of this type. The resulting retrieval is presented back to the user on the web browser as a listing of time-series parameters for the requested well. At LLNL we have extended this simple model in two important ways. The first involves the addition of an alternate method of constraining the initial query by geographic extent; the second entails the addition of optional post-processing of the data retrieved by the query. Both make use of the concept of remote X-clients and require an X-server to be resident on the user's machine.

The data access system at LLNL exploits the web server's ability to launch an application through the common gateway interface (CGI). We make use of a simple CGI compliant, map utility program called HotMap, providing users with the ability to select objects directly from a site map. This simple geographical information system (GIS) runs as a remote X-client and is used to provide an easy method to indicate which wells, boreholes, or piezometers are of interest. The user can point and click to select individual or multiple sampling locations, or can use the mouse to define a rectangular area containing all objects of interest. HotMap also allows many of the functions provided by more expensive GIS systems, including image overlays, panning, zooming, and x-y coordinate indicators. In the example above, the form could be modified to omit selection of a well identifier. The HotMap program replaces the simple web server script that initiated the embedded SQL program. HotMap initiates the embedded SQL program, passing it the sampling location selections made by the user with the mouse.

A similar strategy is used for post-processing of the retrieved data. Data is rarely useful in the native retrieved format. Users frequently spend additional time preparing the data for use in a statistical procedure, scientific visualization or plotting software program. In our example, a common use of time-series parameter data is the creation of a two-dimensional time-series plot. Rather than being presented in tabular form, data resulting from the database query can be passed to a simple X-client plotting program, and then presented graphically as a time-series plot.

These two strategies have been used to provide a rich set of simple, platform independent tools for use by geologists, chemists and other analysts to

provide cost-effective access to information needed for their particular task in the investigation in an immediately useful form. From the desktop, a custom query like the one used in our example, can be formulated and a resulting time-series plot can be produced in less than 10 seconds. Before implementation of this system, traditional retrieval methods might involve two (or more) people and require one or two days before a finished time-series plot could be available.

Work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.